A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR--89-2619

DE89 016604

TITLE Nonlinear Discrete Models for DNA Dynamics

AUTHOR(S) V. Muto
A.C. Scott
P.L. Christiansen
P.S. Lomdahl

SUBMITTED TO To ve prestnted to V International Conference on Waves and Stability in Continuous Media, Sorrento (Italy), October 9-14, 1989.

DISCLAIMER

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

To be presented to
V International Conference on Waves and Stability in Continuous media,
Sorrento (Italy), October 9-14, 1989.

August 1989

Nonlinear discrete models for DNA dynamics

V. Muto*, A.C. Scott**, P.L. Christiansen** and P.S. Lomdahl*

* Center for Nonlinear Studies
and Theoretical Division
Los Alamos National Laboratory
Los Alamos, 87545 New Mexico, USA

** Laboratory of Applied Mathematical Physics
The Technical University of Denmark
DK-2800 Lyngby, Denmark

Abstract

In this paper we report the investigations on a problem related to *DNA* dynamics: thermal generation of localized pulses that are called *solitons*.

The thermal generation of solitons in circular homogeneous *DNA* is investigated by calculating the number of solitons as a function of the absolute temperature. These calculations are effected by using two different models for the *DNA* molecule. In both models the parameters are chosen to match experimentally measured properties of the molecule. We find that a significant number of solitons is generated at physiological temperatures, and a $T^{1/3}$ law is followed at low temperatures.

Introduction

The understanding of energy transport along linear-chain molecules is a long-standing problem that remains of great interest. Molecular dynamics simulations have explored moderately anharmonic motion of short strands of deoxyribonucleic acid (*DNA*) molecule near thermal equilibrium [1], and solitons have been suggested by several researchers [2-7]. In the problem of solitons playing a role in the dynamics of *DNA*, there is a particular aspect related to the question if solitons are generated thermally at physiological temperature. In this work we report on studies performed in order to answer to this question. Two simple nonlinear spring and mass systems are used to model longitudinal wave propagation on a homogeneous circular *DNA* molecule at different temperatures [8-12]. In the first model we consider the *DNA* molecule as a one-dimensional Toda lattice with parameters chosen to match experimentally measured properties [8-11]. The second model is a two-dimensional one [12]. In this case we represent the *DNA* by two Toda chains linked together through springs with restoring force described by a Lennard-Jones potential [13-16]. Again the parameters are chosen to match experimentally measured properties of the molecule.

The one-dimensional model

Here the *DNA* is modeled as a Toda lattice [8-11]. Each particle in the lattice represents a single base pair, and the nonlinear spring represents the Van der Waals potential between adjacent base pairs. Therefore considering a homogeneous system of N masses (each with mass M), whose *longitudinal positions* are denoted by $y_n(t)$, $n = 1, 2, ..., N$, and connected to their nearest neighbours with massless nonlinear springs of potential $V(y_n - y_{n-1})$, Newton's second law for the longitudinal motion of the masses is the set of ordinary differential equations

$$M\ddot{y}_n = V'(y_{n+1} - y_n) - V'(y_n - y_{n-1}), n = 1, 2, ..., N. \quad (1)$$

Here dot denotes time differentiation and prime indicates a derivative with respect to the argument. y_0 and y_{N+1} satisfy the periodicity condition. The exponential potential used by Toda [17] is

$$V(y_{n+1} - y_n) = \frac{a}{b}\exp[-b(y_{n+1} - y_n)] + a(y_{n+1} - y_n) \qquad (2)$$

which has the nature of physical atomic forces for a and b positive parameters. Moreover the Toda lattice is exactly integrable [17], and well developed analytical and numerical techniques can be used to efficiently count the number of solitons. Indeed here the method to count the solitons in a periodic chain is the *spectral analyzer*, based on the inverse spectral theory for the Toda lattice [17]. The spectral analyzer of the Toda lattice, described in detail elsewhere [8,10], has been used through out the computations to count the number of solitons present in the system. And we have assumed that the method still works for the perturbed versions of the Toda lattice.

By fitting the Toda potential to a 6-12 Van der Waals potential and considering experimentally measured properties of *DNA* [8-11], it is found that the mass M of each particle representing a base pair is $M = 1.282 \times 10^{-24} kg$, the nonlinear parameter b is $6.176 \times 10^{10} m^{-1}$ and the parameter a is $5.127 \times 10^{-10} N$.

Two different approaches to the thermalization of the *DNA* molecule have been considered. At first, by assuming the amount of total energy in the system equal to $k_B T$ (here k_B is the Boltzmann constant and T is the absolute temperature measured in K), the system is thermalized by choosing the initial displacement of the masses and their initial velocities from Gaussian random distributions, whose variances are related to the absolute temperature T [8-10]. It is found that a significant number of solitons, N_S, is generated at physiological temperature. Moreover, at low temperatures, a $T^{1/3}$ law is found for the dependence of the number of solitons on the temperature [8-11]. A typical curve is shown in figure 1 where $\log(N_S/N)$ is plotted versus $\log(T)$. The full line is the curve computed for a normalized version of the unperturbed equations (1), for $N = 65000$ and with initial conditions such that the total energy is only kinetic energy. At physiological temperature, $310K$, the ratio N_S/N has a value approximately equal to 0.31.

In the second approach to thermalization, in order to describe the interaction of the system with a thermal reservoir at a finite temperature, a damping and a noise force which simulate a thermal bath are added to the equations of motion for the molecular displacement, thereby obtaining Langevin equations [18,19]. Then [10,11] the equation of motion of each particle is given by the perturbed Toda equation

$$M\ddot{y}_n = V'(y_{n+1} - y_n) - V'(y_n - y_{n-1}) - M\Gamma\dot{y}_n + \eta_n(t), \qquad (3)$$

where Γ is the damping coefficient and $\eta_n(t)$ is the random force, assumed to be Gaussian process with zero mean and correlation function of the form

$$< \eta_n(t)\eta_{n'}(t') >= 2M\Gamma k_B T \delta_{nn'}\delta(t-t'). \tag{4}$$

Here $\delta_{nn'}$ is the Kronecker delta symbol and $\delta(t-t')$ is the Dirac delta function.

Numerical integration of a normalized version of the perturbed equations of motion (3), with different values for the normalized damping coefficient α, given by $\alpha = \Gamma\sqrt{\frac{M}{a\delta}}$, shows that, when thermal equilibrium is reached, in a system of $N = 32$ masses, the same significant number of solitons is generated in *DNA* at physiological temperature [10,11], namely $N_S/N \simeq 0.31$.

Moreover from dynamical simulations for a system with $N = 32$, at different temperatures T, and at various α's it is found a good agreement with the $T^{1/3}$ law obtained as dependence of the number of solitons on the temperature [10,11]. A typical case is shown in figure 1 where $\log(N_S/N)$ is plotted versus $\log(T)$. The values denoted by the circle, the square and the triangle have been calculated for the normalized version of the perturbed equations (3), with a number of masses $N = 32$ and normalized damping coefficients $\alpha = 5.0$, $\alpha = 3.0$ and $\alpha = 0.85$, respectively.

Moreover we note that a study on how the number of solitons depends on the number of masses in this one-dimensional system has been considered in [11].

The two-dimensional model

Here the *DNA* is described by two chains transversally coupled, where each chain simulate one of the two polynucleotide strands of the *DNA* molecule [12]. Each of the two chains is a spring and mass system. Each mass of the model represents a single base of the base pair. As in the above mentioned one-dimensional model, the longitudinal springs, connecting masses of the same strand, represent the Van der Waals potential between adjacent base pairs. The transverse springs, connecting corresponding masses of the two strands, represent the hydrogen bonds that connect the two bases in a pair. Again we assume to deal with a homogeneous *DNA* molecule, therefore each particle in our mass and spring system has mass M, and the springs are assumed to be massless.

For each base pair, the model includes four degrees of freedom, u_n, x_n, and v_n, y_n, for the two strands, respectively. The $u_n = u_n(t)$ and $v_n = v_n(t)$,

$n = 1, 2, ..., N$ denote the *transverse diplacements*, namely they correspond to the displacements of the bases from their equilibrium positions along the direction of the hydrogen bonds that connect the two bases of the base pair. The $x_n = x_n(t)$ and $y_n = y_n(t)$, $n = 1, 2, ..., N$ denote the *longitudinal diplacements*, namely they correspond to the displacements of the bases from their equilibrium positions along the direction of the phosphodiester bridge that connect the two bases of the same strand.

As in the previous model, the anharmonic Van der Waals potentials are described by the Toda potential. Denoting by I and II the two strands, the anharmonic potentials are given by

$$V_{T,I}(l_n' - l_L) = \frac{\tilde{a}}{\tilde{b}} \exp[-\tilde{b}(l_n' - l_L)] + \tilde{a}(l_n' - l_L) , \tag{5}$$

and

$$V_{T,II}(l_n'' - l_L) = \frac{\tilde{a}}{\tilde{b}} \exp[-\tilde{b}(l_n'' - l_L)] + \tilde{a}(l_n'' - l_L) , \tag{6}$$

for the two sdrands, respectively.

In expressions (5) and (6) $\tilde{a}$ and $\tilde{b}$ are parameters which will be given later. Moreover in the above expressions l_L is the equilibrium distance between two bases in the same strand, namely it is the distance along the helix axis between adjacent base pairs, ($l_L = 3.4$Å [20]). Finally l_n' denotes the distance between two bases in the strand denoted by I, and similarly for l_n'' in the strand denoted by II. Their expressions are given by $l_n' = \sqrt{(l_L + x_{n+1} - x_n)^2 + (u_{n+1} - u_n)^2}$ and $l_n'' = \sqrt{(l_L + y_{n+1} - y_n)^2 + (v_{n+1} - v_n)^2}$, respectively.

The two bases in a pair are connected through hydrogen bonds which we model by a Lennard-Jones potential. This type of potential is generally accepted to model the effect of hydrogen bonds [13-16]. Thus the anharmonic potential for the transverse springs modeling the hydrogen bonds is given by

$$V_{LJ}(l_n' - l_T + l_H) = 4\,\varepsilon \left[\left(\frac{\sigma}{l_n' - l_T + l_H} \right)^{12} - \left(\frac{\sigma}{l_n' - l_T + l_H} \right)^{6} \right] . \tag{7}$$

In expression (7) ε and σ are parameters which will be given later. The lenght of the hydrogen bond between the two bases in a pair is given by $l_n' - l_T + l_H$.

Here $t_n^{'}$ denotes the distance between two bases of the two strands and its expression is given by $t_n^{'} = \sqrt{(l_T + v_n - u_n)^2 + (y_n - x_n)^2}$. Moreover l_T is the equilibrium distance between adjacent bases in a pair, namely the diameter of the helix, ($l_T = 20\text{Å}$ [20]), and $l_H = 2^{1/6}\sigma$ is the equilibrium length of the hydrogen bond.

As above, by fitting the Toda potential to a 6-12 Van der Waals potential and considering experimentally measured properties of *DNA*, it is found [12] that the mass $\tilde{M}$ of each base is $\tilde{M} = 6.41 \times 10^{-25}\ kg$, the nonlinear parameter $\tilde{b}$ is $6.176 \times 10^{10} m^{-1}$ and the parameter $\tilde{a}$ is $2.5635 \times 10^{-10} N$. And for the Lennard-Jones' parameters, the values $\varepsilon = \varepsilon_{LJ} = 0.22\ eV = 0.35244 \times 10^{-19}\ Nm$ and $\sigma = 4.01 \times 10^{-10}\ m$ can be considered [12].

The equations of motion for the system described above are obtained from the Hamiltonian of the system

$$H = \sum_{n=1}^{N} \left[T_n + V_{T,I}(l_n^{'} - l_L) + V_{T,II}(l_n^{''} - l_L) + V_{LJ}(t_n^{'} - l_T + l_H) \right] , \qquad (8)$$

where N is the number of base pairs in the *DNA* molecule. The kinetic energy T_n is given by

$$T_n = \frac{1}{2}\tilde{M}(\dot{x}_n^2 + \dot{u}_n^2) + \frac{1}{2}\tilde{M}(\dot{y}_n^2 + \dot{v}_n^2) , \qquad (9)$$

and the anharmonic potentials $V_{T,I}(l_n^{'} - l_L)$, $V_{T,II}(l_n^{''} - l_L)$ and $V_{LJ}(t_n^{'} - l_T + l_H)$ are given by expressions (5), (6) and (7), respectively.

In order to describe the interaction of the system with a thermal reservoir at a finite temperature, a damping force and a noise force are added to the equations of motion for the molecular displacements [18,19].

If we denote by $d_n^{(i)}$, $i = 1,2,3,4$ and $n = 1,2,...,N$, the displacement variables, namely $d_n^{(1)} = x_n$, $d_n^{(2)} = u_n$, $d_n^{(3)} = y_n$, $d_n^{(4)} = v_n$, the equations of motion derived from the Hamiltonian (8) can be briefly written as

$$\tilde{M}\ddot{d}_n^{(i)} = -\frac{\partial H}{\partial d_n^{(i)}} , \qquad (10)$$

where i = 1,2,3,4, and n = 1,2,...,N.

Therefore to describe the interaction of the system with a thermal reservoir at temperature T, a damping force and a noise force are added to eqs. (10), namely

$$F_n^{(i)} = -\tilde{M}\Gamma\ \dot{d}_n^{(i)} + \eta_n^{(i)}(t)\ . \tag{11}$$

Here Γ is the damping coefficient and $\eta_n^{(i)}(t)$, $i = 1,2,3,4$, are the random forces acting on the bases. For the noises $\eta_n^{(i)}(t)$ it is assumed that they are gaussian processes with zero mean and that their correlation function has the form

$$\left\langle \eta_n^{(i)}(t)\ \eta_{n'}^{(j)}(t') \right\rangle\ =\ \tilde{M}\Gamma k_B T\ \delta_{nn'}\ \delta_{ij}\ \delta(t-t')\ . \tag{12}$$

Then the equations of motion for the system of particles are given by the perturbed version of system (10), namely

$$\tilde{M}\ddot{d}_n^{(i)} = -\frac{\partial H}{\partial d_n^{(i)}} - \tilde{M}\Gamma\ \dot{d}_n^{(i)} + \eta_n^{(i)}(t)\ , \tag{13}$$

where i = 1,2,3,4, and n = 1,2,...,N.

In order to compute the number of solitons (N_S) in the double chain, we have averaged on the values for the number of solitons in each one of the two chains. The latter values have been computed using again the spectral analyzer of the Toda lattice. Our typical results are shown in figure 2, where $\log(N_S/N)$ is plotted versus $\log(T)$ for different choices of the Lennard-Jones parameter ϵ. In particular the perturbed system (13) chosen as model of the double chain immerged in a thermal bath has been integrated for a number of masses $N = 32$ for each of the two identical chains, and with a damping coefficient $\alpha = 1.0$. The values denoted by the square are computed for $\epsilon = \epsilon_{LJ}/10$, and the ones denoted by diamond are computed for $\epsilon = \epsilon_{LJ}/5$. The other parameters in system (13) are the ones considered for the *DNA* molecule.

In figure 2 the present results are compared with the ones obtained with the one-dimensional model. In particular the full line describes the results obtained using the simple unperturbed Toda lattice. And the values denoted by the cross are the results obtained considering the perturbed model, with a damping coefficient $\alpha = 1.0$ and a number of masses $N = 32$.

These results show that, when thermal equilibrium is reached, in a chain of $N = 32$ masses, approximately the same significant number of solitons is generated in

DNA at physiological temperature, namely $N_S/N \simeq 0.335$. Moreover from dynamical simulations for a system with $N = 32$ and at different temperatures T, it is found a good agreement with the $T^{1/3}$ law obtained as dependence of the number of solitons on the temperature. again for temperatures less than $100K$ [8-12].

Finally it is important to note that this two-dimensional system has been considered [12] as model for the denaturation of the *DNA* molecule, with particular attention to the phenomenon known as *melting*, that is when the two sdrands of the *DNA* helix readly come apart.

Conclusions

In this work we have presented a study on the thermal generation of solitons in *DNA*, considering two different models for the *DNA* dynamics.

The first model is a one-dimensional representation of *DNA*. The molecule is thought as a simple one-dimensional system of masses and spring, with restoring forces obtained by the exponential Toda potential. Each mass in the system represent a base pair.

The second model is a two-dimensional representation of the molecule. In this case the *DNA* is thought as two chains of masses and springs, with exponential Toda restoring forces, linked transversally together through springs with restoring force described by a Lennard-Jones potential. Each mass in the system represent a base of a pair of the molecule.

With the one-dimensional model, two different approaches to thermalization have been considered. The first approach is realized by choosing random initial distributions for the displacements and the velocities of the masses such that the total energy in the system is $k_B T$. On the other hand the second approach to thermalization is performed perturbing the equations of motion with a damping force and a noise force to model the interaction of the *DNA* molecule with a thermal reservoir at temperature T. Only this second approach to thermalization is used for the two-dimensional model.

Using these models we have analyzed the dynamics of a short homogeneous circular *DNA* molecule of 32 base pairs. In both cases it is found that a significant number of solitons, more than 1/3 of the number of base pairs, is generated at physiological temperature. Moreover we have observed that the number of solitons follows the power law $N_S \propto T^{1/3}$, for a wide temperature range, namely for $T < 100K$.

Figure Captions

Figure 1. Logarithm to base 10 of the ratio N_S/N versus the logarithm to base 10 of the temperature measured in K. The full line is the curve calculated for the unperturbed equations (1), with a number of masses in the Toda chain $N = 65000$, and initial conditions such that the total energy is only kinetic energy. The values denoted by the circle, the square and the triangle are calculated for the perturbed equations (3) for the value of the normalized damping coefficients $\alpha = 5.0$, $\alpha = 3.0$ and $\alpha = 0.85$, respectively, and with a number of masses $N = 32$.

Figure 2. Logarithm to base 10 of the ratio N_S/N versus the logarithm to base 10 of the temperature T measured in K. The full line is the curve calculated for the unperturbed (one-dimensional) equations (1), with a number of masses in the Toda chain $N = 65000$, and initial conditions such that the total energy is only kinetic energy. The values denoted by the crosses are the ones obtained from the perturbed (one-dimensional) system (3) for a damping coefficient $\alpha = 1.0$ and a number of masses $N = 32$. The values denoted by the square and the diamond are calculated for the perturbed (two-dimensional) system (13) with *DNA* parameters, and Lennard-Jones parameter ε equal to $\varepsilon = \varepsilon_{LJ}/10$ and $\varepsilon = \varepsilon_{LJ}/5$, respectively. Damping coefficients $\alpha = 1.0$, and $N = 32$ masses in each chain.

References

[1] M.J. Levitt, Cold Spring Harb. Sym. Quan. Biol. **47**, 251 (1983).

[2] S.W. Englander, N.R. Kallenbach, A.J. Heeger, J.A. Krumhansl and S. Litwin, Proc. Natl. Acad. Sci. (USA) **77** (12), 7222 (1980).

[3] S. Yomosa, J. Phys. Soc. Japan **51**, 3318 (1982).

[4] S. Takeno and S. Homma, Prog. Theor. Phys. **70**, 308 (1983).

[5] A. Banerjee and H.M. Sobell, J. Biomol. Struct. Dyn. **1**, 253 (1983).

[6] A.C. Scott, Phys. Rev. A **31**, 3518 (1985).

[7] V. Muto, J. Halding, P.L. Christiansen and A.C. Scott, J. Biomol. Struct. Dyn. **5**, 873 (1988).

[8] V. Muto, *Nonlinear Models for DNA Dynamics*, Ph.D. Thesis, The Danish Center for Applied Mathematics and Mechanics, report N. S 47, October 1988.

[9] V. Muto, A.C. Scott and P.L. Christiansen, Phys. Lett. A **136** (1,2), 33 (1989).

[10] V. Muto, A.C. Scott and P.L. Christiansen, *A Toda lattice model for DNA: thermally generated solitons*, submitted for publication on Physical Review B.

[11] V. Muto, A.C. Scott and P.L. Christiansen, *A nonlinear model for DNA dynamics*, submitted for publication to the Proceedings of the "VII International Conference on Mathematical and Computer Modelling", Chicago (Illinois), August 2-5, 1989.

[12] V. Muto, P.S. Lomdahl and P.L. Christiansen, *A two-dimensional discrete model for DNA*, in preparation.

[13] M.J. Levitt, J. Mol. Biol. **168**, 595 (1983).

[14] J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acid*, Cambridge University Press, Cambridge, (1987).

[15] P. Perez and N. Theodorakopoulos, Phys. Lett. **117 A**, 405 (1986); Phys. Lett. **124 A**, 267 (1987).

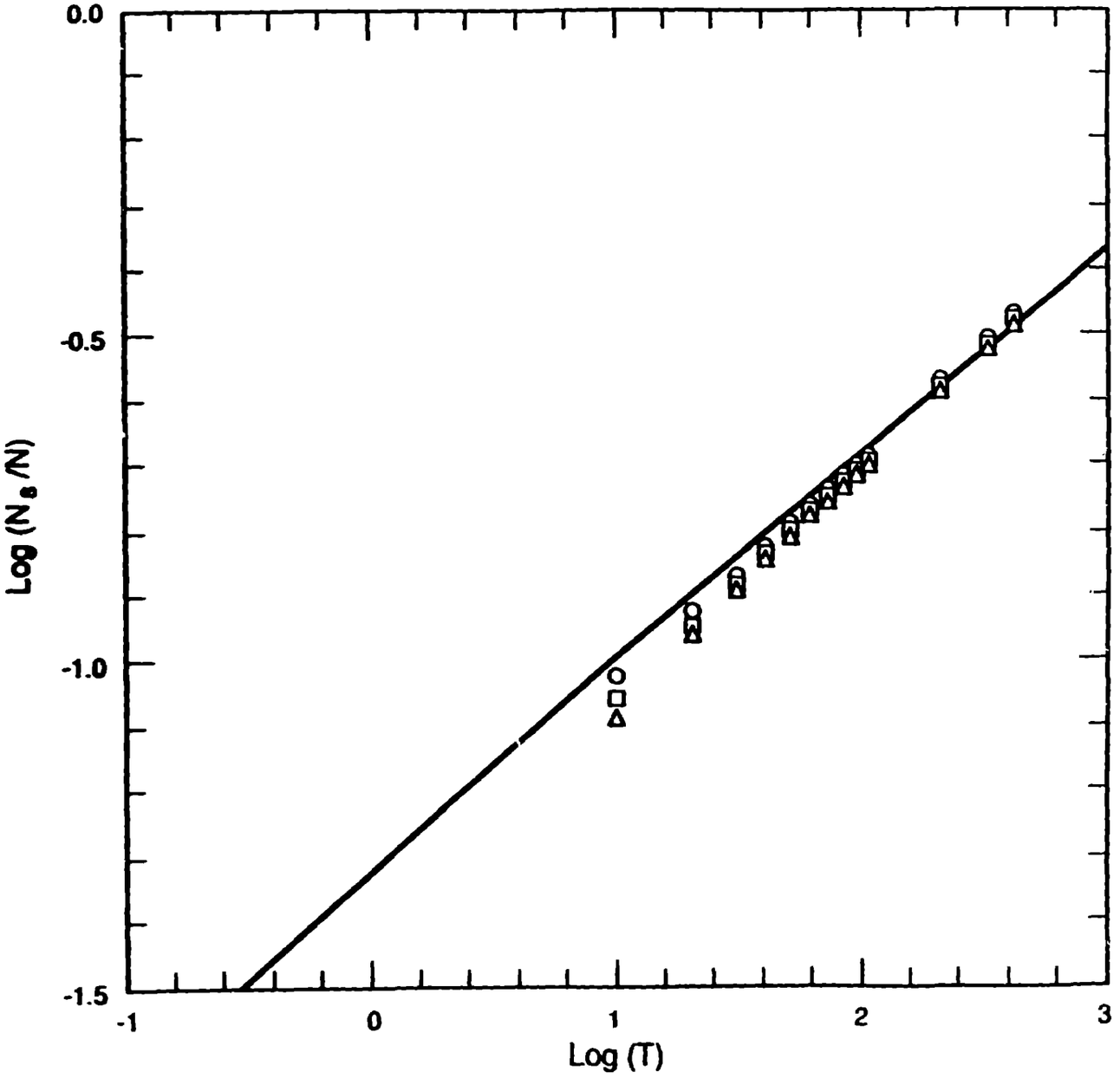
[16] J. Halding and P.S. Lomdahl, Phys. Lett. **124 A**, 37 (1987).

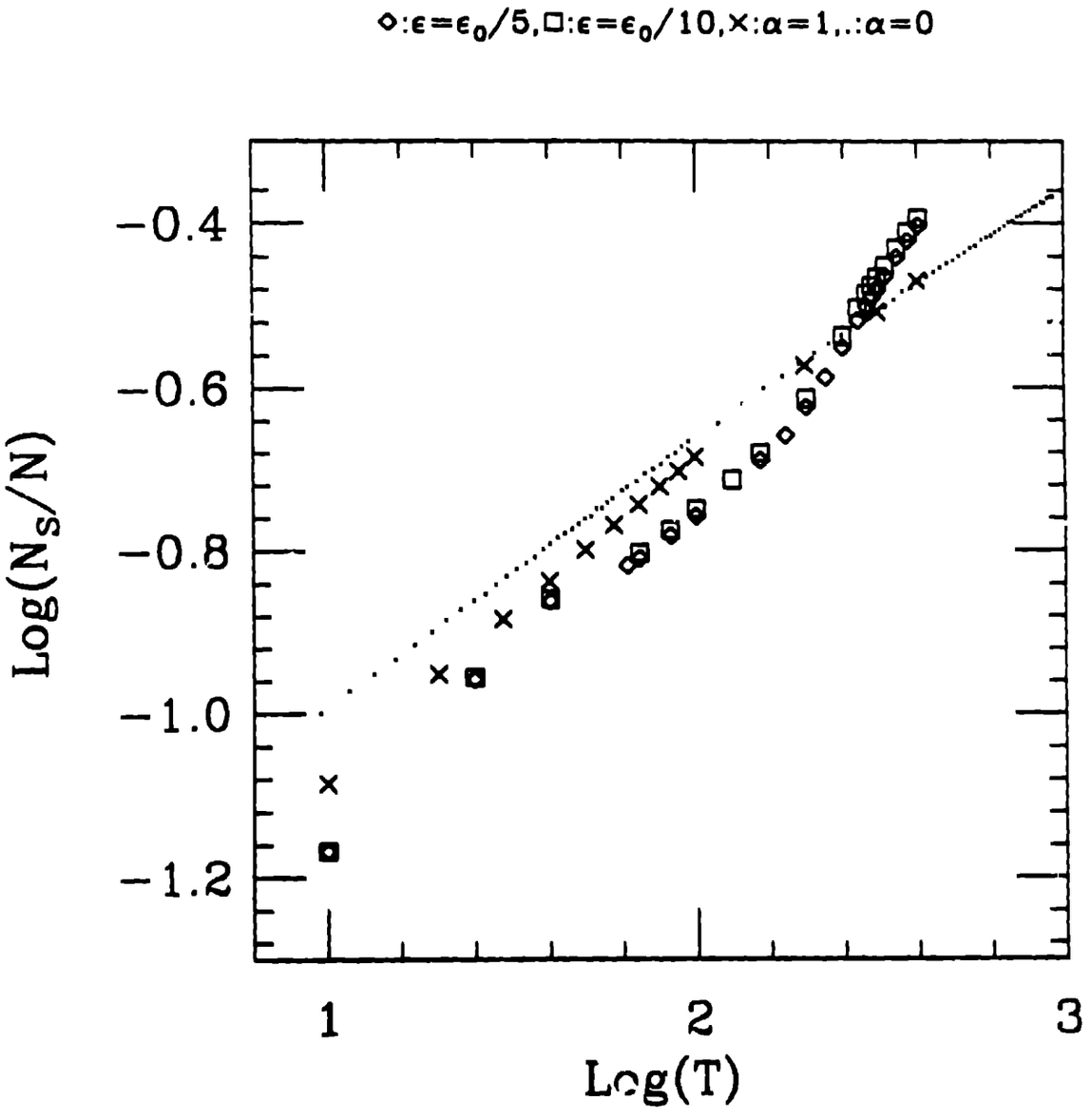
[17] M. Toda, *Theory for Nonlinear Lattices*, Springer Series in Solid State Sciences, 20, Springer-Verlag (1981).

[18] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[19] H. Risken, *The Fokker-Plank equation, Methods of Solution and Applications*, Springer, Berlin, (1984).

[20] L. Stryer, *Biochemistry*, W.H. Freeman and Co., (1981).


0.0
-0.5
-1.0
-1.5
-1
0
1
2
3
Log (N$_s$/N)
Log (T)


◇:ε=ε0/5,□:ε=ε0/10,×:α=1,.:α=0
−0.4
−0.6
−0.8
−1.0
−1.2
Log(Ns/N)
1
2
3
Log(T)